Appl. Math. J. Chinese Univ.
2016, 31(2): 127-147

Panel data models with cross-sectional dependence: a selective review

XU Qiu-hua[1] CAI Zongwu[2,1] FANG Ying[1,*]

Abstract. In this review, we highlight some recent methodological and theoretical developments in estimation and testing of large panel data models with cross-sectional dependence. The paper begins with a discussion of issues of cross-sectional dependence, and introduces the concepts of weak and strong cross-sectional dependence. Then, the main attention is primarily paid to spatial and factor approaches for modeling cross-sectional dependence for both linear and nonlinear (nonparametric and semiparametric) panel data models. Finally, we conclude with some speculations on future research directions.

§1 Introduction

The analysis of panel (longitudinal) data has attracted considerable attention in many applied fields as economics and finance as well as biomedicine during the last three decades. Panel data models provide researchers with multiple observations on each individual considered in a given sample and thus have the ability to control unobserved heterogeneity and uncover dynamic relationships that can not be obtained by using either pure cross-sectional or time series data. Baltagi (2013) and Hsiao (2014) provide comprehensive surveys to the existing models for panel data analysis. However, most of the existing models in the literature assume that there is no correlation among cross-sectional units and it might not be appropriate for many real applications. For example, in empirical studies using panel data sets of regions, states or countries, the cross-sectional units could be interdependent (termed as cross-sectional dependence) due to competition, spill-overs, externalities, etc. Moreover, theoretically, estimators obtained by ignoring cross-sectional dependence could be inconsistent. These facts prompt the swift growing demand for modeling cross-sectional dependence in both theoretical and methodological research and real applications.

Received: 2016-01-31.
MR Subject Classification: 62G02, 62J02.
Keywords: Panel data models, cross-sectional dependence, spatial dependence, interactive fixed effects, common factors.
Digital Object Identifier(DOI): 10.1007/s11766-016-3441-9.
Supported by the National Natural Science Foundation of China (71131008(Key Project) and 71271179).
* Correspondence author.

As pointed out by Chudik et al. (2011), characterizing cross-sectional dependence for panel data lies on the size of the time series dimension (T) of the panel relative to its cross-sectional dimension (N). When N is fixed and T is large, cross-sectional dependence can be modeled using the so-called seemingly unrelated regression method (SUR) as in Zellner (1962). However, when N is large relative to T, the SUR method is no longer feasible. There are two major methods proposed recently in the literature: spatial and factor approaches. Contrary to time series analysis, where the past and distant future are always assumed to be asymptotically independent, there is no such natural ordering among different cross-sectional units. Consequently, to deal with cross-sectional dependence, a correlation structure has to be imposed. For example, in spatial econometrics, the cross-sectional correlation structure is modeled through a pre-specified spatial weighting matrix which often depends on the geographic locations of cross-sectional units or some more general economic variables. In panel data models with factor structure, the correlation among cross-sectional individuals is introduced by a finite number of unobserved common factors that influence each individual.

In this paper, we aim to provide a resource that surveys the state of art in estimation and testing of panel data models with cross-sectional dependence for both linear and nonlinear models. Indeed, Sarafidis and Wansbeek (2012) provide an excellent survey on linear panel data models with cross-sectional dependence. Differently from Sarafidis and Wansbeek (2012), we include some more recent developments in this area and in particular, we provide an overview of nonparametric and semiparametric estimation and testing of panel data models with cross-sectional dependence, which are not covered in Sarafidis and Wansbeek (2012).

The rest of the paper is organized as follows. In Section 2 we discuss potential problems when ignoring cross-sectional dependence and then introduce the concepts of weak and strong cross-sectional dependence. Section 3 reviews the various estimation methods available for linear panel data models with cross-sectional dependence. Section 4 focuses on the estimation and testing of nonparametric and semiparametric panel data models with cross-sectional dependence. Section 5 concludes with discussions on some open and interesting research problems.

§2 Weak and strong cross-sectional dependence

2.1 Inconsistency when ignoring cross-sectional dependence

There have been many studies in the literature on the impact of ignoring cross-sectional dependence on panel data regression models; see, for example, Phillips and Sul (2007), Hsiao and Tahmiscioglu (2008) and the references therein. In this subsection, we illustrate this issue by recalling one of the examples used in Phillips and Sul (2007).

Consider the following linear dynamic panel data model:

$$y_{it} = \rho y_{i,t-1} + e_{it}, \quad |\rho| < 1 \quad \text{with} \quad e_{it} = \gamma_i f_t + \varepsilon_{it}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{1}$$

where y_{it} is the observed dependent variable on the ith cross-sectional unit at time t, the error e_{it} in (1) has a single factor (f_t) structure, γ_i is the factor loading for unit i, and ε_{it} is the idiosyncratic error of y_{it}, which is assumed to be $\mathrm{IID}(0,\sigma^2)$. Model (1) can be rewritten as

$$y_{it} = y^0_{it} + \gamma_i G_t, \quad y^0_{it} = \rho y^0_{i,t-1} + \varepsilon_{it}, \quad G_t = \rho G_{t-1} + f_t.$$

Phillips and Sul (2007) show that the probability limit of the pooled least squares estimate is

$$\operatorname*{plim}_{N\to\infty}(\hat\rho - \rho) = \frac{\operatorname*{plim}_{N\to\infty}(1/N)\sum_{i=1}^N\sum_{t=1}^T y_{i,t-1}e_{it}}{\operatorname*{plim}_{N\to\infty}(1/N)\sum_{i=1}^N\sum_{t=1}^T y^2_{i,t-1}} = \frac{m^2_\gamma(\sum_{t=1}^T G_{t-1}f_t)}{T[\sigma^2/(1-\rho^2)] + m^2_\gamma\sum_{t=1}^T G^2_{t-1}}, \tag{2}$$

where $m^2_\gamma \equiv \lim_{N\to\infty}(1/N)\sum_{i=1}^N \gamma_i^2$ is assumed to be finite. In view of (2), it is evident to see that the pooled least squares estimator, $\hat\rho$, which ignores the presence of cross-sectional dependence, is inconsistent and it converges to a random variable rather than a constant when T is fixed and $N\to\infty$. Moreover, the inconsistency of $\hat\rho$ depends on the degree of cross-sectional dependence. In fact, Hsiao (2014) points out that the bias of panel estimators could vanish when the degree of cross-sectional dependence is weak. More specifically, let $\mathbf{e}_t = (e_{1t},\dots,e_{Nt})'$ be the $N\times 1$ vector which stacks the errors of N cross-sectional units at time t and $\boldsymbol{\Sigma}$ be its covariance matrix which is $N\times N$. If the number of nonzero elements in each row of $\boldsymbol{\Sigma}$ is bounded by h_N and $h_N/N\to 0$ as $N\to\infty$, estimators which ignore cross-sectional dependence could still be consistent when T is finite. However, if h_N/N converges to a nonzero constant as $N\to\infty$, the asymptotic bias of estimators which ignore cross-sectional dependence could be random as shown in (2) no matter how large N is when T is finite. The above two cases correspond to the notions of weak and strong cross-sectional dependence in the literature which we will discuss in the next subsection.

2.2 Weak and strong cross-sectional dependence

Since the degree of cross-sectional dependence has a major impact on estimation of panel data models, many researchers have proposed some methods to define weak and strong cross-sectional dependence. For example, Forni and Lippi (2001) and Deistler et al. (2010) define these concepts for covariance stationary processes. In this subsection, we introduce more general definitions for these concepts given by Chudik et al. (2011) which do not assume covariance stationary for the underlying processes.

Let $\{e_{it}, 1\le i\le N, 1\le t\le T\}$ be any double index process and $\mathcal{I}_t$ be the information set available at time t. For each t, suppose that $E(\mathbf{e}_t|\mathcal{I}_{t-1}) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{e}_t|\mathcal{I}_{t-1}) = \boldsymbol{\Sigma}_t$, where $\boldsymbol{\Sigma}_t$ is an $N\times N$ non-negative definite matrix whose elements are uniformly bounded. In addition, let $\mathbf{w}_t = (w_{1t},\dots,w_{Nt})', 1\le i\le N, 1\le t\le T$, be a non-stochastic weight vector. For any t, $\{\mathbf{w}_t\}$ satisfies the following *granularity conditions* as $N\to\infty$:

$$\|\mathbf{w}_t\| = O(N^{-1/2}), \tag{3}$$

$$\frac{w_{jt}}{\|\mathbf{w}_t\|} = O(N^{-1/2}) \quad \text{for any } j,\ 1\le j\le N, \tag{4}$$

where $\|\cdot\|$ denotes the Euclidean norm.

Definition 2.1. (Weak and strong cross-sectional dependence, Chudik et al. (2011)). The process $\{e_{it}, 1\le i\le N, 1\le t\le T\}$ is said to be weakly dependent at a given point in time t

conditional on $\mathcal{I}_{t-1}$, if for any sequence of weight vectors $\{\mathbf{w}_t\}$ satisfying (3) and (4) we have

$$\lim_{N\to\infty} \mathrm{Var}(\mathbf{w}_t'\mathbf{e}_t|\mathcal{I}_{t-1}) = 0.$$

$\{e_{it}, 1 \le i \le N, 1 \le t \le T\}$ is said to be strongly dependent at a given point in time t conditional on $\mathcal{I}_{t-1}$, if there exists a sequence of weight vectors $\{\mathbf{w}_t\}$ satisfying (3) and (4) and a constant K independent of N such that for any N sufficiently large

$$\mathrm{Var}(\mathbf{w}_t'\mathbf{e}_t|\mathcal{I}_{t-1}) \ge K > 0$$

as $N \to \infty$.

Let $\lambda_{max}(A)$ be the maximum eigenvalue of A. Chudik et al. (2011) further show that (i) $\{e_{it}\}$ is weakly dependent at a point in time t, if $\lambda_{max}(\mathbf{\Sigma}_t)$ is bounded in N, and (ii) $\{e_{it}\}$ is strongly dependent at a point in time t, if and only if for any N sufficiently large, $N^{-1}\lambda_{max}(\mathbf{\Sigma}_t) \ge K > 0$ as $N \to \infty$.

In the following section, we will see that the spatial correlation and the cross-sectional dependence introduced by factor structure are weakly and strongly dependent, respectively.

§3 Linear panel data models with cross-sectional dependence

In this section, we survey the existing literature for estimation of linear panel data models with cross-sectional dependence. We first focus on the SUR method which is appropriate for small N and large T panels. Then we discuss the spatial and factor approaches when the SUR method can not be used.

3.1 SUR approach

When N is fixed and $T \to \infty$, cross-sectional dependence can be modeled using the SUR approach proposed by Zellner (1962). Specifically, consider the following model

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\beta_i + e_{it},$$

where α_i is an individual-specific effect, β_i is a $p \times 1$ vector of unknown coefficients, and $\mathbf{x}_{it}$ is a $p \times 1$ vector of explanatory variables on the ith cross-sectional unit at time t. The regressors satisfy the strongly exogenous condition; that is, $E(e_{it}|\mathbf{X}_1, \ldots, \mathbf{X}_T) = 0$, where $\mathbf{X}_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{Nt})'$. The SUR approach results in a feasible generalized least squares (FGLS) estimator in which OLS is conducted to each individual-specific equation to get consistent estimators of $\{\beta_i\}_{i=1}^N$ at the first stage. The estimators of $\{\beta_i\}_{i=1}^N$ are then used to compute the residuals $\{\hat{e}_{it}\}_{1\le i\le N, 1\le t\le T}$ which are employed to estimate the covariance between units i and j using $\frac{1}{T}\sum_{t=1}^T \hat{e}_{it}\hat{e}_{jt}$. At the second stage, the coefficient estimators are obtained using generalized least squares with the inverse of the estimated covariance matrix as a weighting matrix. The iterative procedure may then be continued until the coefficient estimators converge.

3.2 Spatial approaches

When $N > T$, it is well known that the SUR approach can not be used. One of the major approaches to modeling cross-sectional dependence in panel data models is the spatial method, which was developed mainly for cross-sectional regression models; see, for example, Anselin (1988), Kelejian and Prucha (1999) and Anselin et al. (2008). The correlation among different individuals in the spatial approach is captured by means of a so-called spatial weights matrix, $W = (w_{ij})_{N\times N}$, whose specification is often ad hoc. According to the "first law of geography" of Tobler (1970), that is, "Everything is related to everything else, but near things are more related than distant things", in practice, the spatial weights matrix is typically pre-specified by geographical factors such as contiguity or distance. In economic applications, more general "economic distance" are often adopted to set the spatial weights matrix as well (see Conley (1999) and Conley and Topa (2002)). By convention, $w_{ii} = 0$ for all i, which indicates that the "distance" between unit i and itself is zero. Moreover, W is often row-normalized so that the sum of each row is 1. In general, W satisfies the following uniformly boundedness condition (see Kapoor et al. (2007) and Lee (2007))

$$\max_{i=1,\ldots,N}\sum_{j=1}^{N}|w_{ij}| \leq M \quad \text{and} \quad \max_{j=1,\ldots,N}\sum_{i=1}^{N}|w_{ij}| \leq M \quad \text{for some constant } M > 0. \tag{5}$$

The dependence relation among cross-sectional units captured by W may pertain to different components in regression models, such as the dependent variable, the explanatory variables and the error terms, via the so-called spatial lag operator which creates a weighted average of the neighboring observations. According to the specific form of the spatial variables, two classes of specifications for spatial panel models can be distinguished: the spatial error panel models (SEM) and the spatial autoregression panel models (SAR).

3.2.1 Spatial error panel models

According to Elhorst (2003), a spatial error panel model can be specified as follows:

$$\mathbf{y}_t = \mathbf{X}_t\beta + \alpha + \phi_t,\ \phi_t = \rho W\phi_t + \varepsilon_t,\ E(\varepsilon_t) = \mathbf{0},\ E(\varepsilon_t\varepsilon_t') = \sigma^2 I_N,$$

where $\mathbf{y}_t = (y_{1t},\ldots,y_{Nt})'$, $\mathbf{X}_t = (\mathbf{x}_{1t},\ldots,\mathbf{x}_{Nt})'$, ϕ_t and ε_t can be defined accordingly, $\alpha = (\alpha_1,\ldots,\alpha_N)'$ is the vector of individual-specific effects, I_N is the identity matrix of dimension N and ρ is the spatial autocorrelation coefficient. Thus, a spatial error panel model applies the spatial lag operator to the error terms. If condition (5) holds, it is obvious that the largest eigenvalue of the covariance matrix of ϕ_t is bounded in N, which implies that the cross-sectional dependence among error terms is weak.

If elements in α are treated as fixed effects, the SEM model with fixed effects can be estimated by maximum likelihood estimation (MLE). First, $\mathbf{y}_t$ and $\mathbf{X}_t$ are demeaned to get rid of the fixed effects. Then, the log-likelihood function is set up as follows

$$\ln L(\mathbf{y}|\rho,\sigma^2,\beta) = -\frac{NT}{2}\ln(2\pi\sigma^2) + T\sum_{i=1}^{N}\ln(1-\rho\lambda_i) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{\varepsilon}_t'\tilde{\varepsilon}_t, \tag{6}$$

where $\tilde{\varepsilon}_t = (I_N - \rho W)[\mathbf{y}_t - \bar{\mathbf{y}} - (\mathbf{X}_t - \bar{\mathbf{X}})\beta]$, $\bar{\mathbf{y}} = (\mathbf{y}_{1.}, \dots, \mathbf{y}_{N.})'$, $\bar{\mathbf{X}} = (\bar{\mathbf{X}}'_{1.}, \dots, \bar{\mathbf{X}}'_{N.})'$, $\mathbf{y}_{i.} = \frac{1}{T}\sum_{t=1}^T y_{it}$, $\bar{\mathbf{X}}_{i.} = \frac{1}{T}\sum_{t=1}^T \mathbf{x}_{it}$ for all i, and $\{\lambda_i\}_{i=1}^N$ are the eigenvalues of W. An iterative two-stage procedure has to be used to maximize (6), since at the first stage, the closed-form expressions of $\hat{\beta}(\rho)$ and $\hat{\sigma}^2(\rho)$ can not be derived; see, for example, Elhorst (2010) for details.

3.2.2 Spatial autoregression panel models

A spatial autoregression panel model incorporates a spatially lagged dependent variable as an explanatory variable in the regression specification and it can be rewritten as

$$\mathbf{y}_t = \delta W \mathbf{y}_t + \mathbf{X}_t\beta + \alpha + \varepsilon_t,\ E(\varepsilon_t) = \mathbf{0},\ E(\varepsilon_t\varepsilon_t') = \sigma^2 I_N,$$

where δ is often called the spatial autoregressive coefficient.[1]

Similar to SEM models, the MLE can be applied to estimating SAR models with fixed effects. Consider the following log-likelihood function

$$\ln L(\mathbf{y}|\delta, \sigma^2, \beta) = -\frac{NT}{2}\ln(2\pi\sigma^2) + T\sum_{i=1}^N \ln(1-\delta\lambda_i) - \frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{\varepsilon}_t'\tilde{\varepsilon}_t, \tag{7}$$

where $\tilde{\varepsilon}_t = (I_N - \delta W)(\mathbf{y}_t - \bar{\mathbf{y}}) - (\mathbf{X}_t - \bar{\mathbf{X}})\beta$. A simple two-stage procedure suffices for estimating the unknown parameters. Specifically, in the first step, the closed-form expressions of $\hat{\beta}(\delta)$ and $\hat{\sigma}^2(\delta)$ can be obtained by maximizing (7) with respect to β and σ^2. In the second step, we substitute $\hat{\beta}(\delta)$ and $\hat{\sigma}^2(\delta)$ into (7) and get the concentrated log-likelihood function. We then maximize this function with respect to δ which yields the MLE estimator of δ.

The reader is referred to the papers by Elhorst (2003, 2010) for estimation of SEM and SAR models with random effects.

3.2.3 Further developments

If the number of cross-sectional units is large, the MLE involves substantial computational issues. Kapoor et al. (2007) introduce a generalized moments (GM) procedure and develop an FGLS estimator for a SEM panel model with error components disturbances. They show that the FGLS estimator is consistent and asymptotic normality when T is fixed and $N \to \infty$.

Lee and Yu (2010a) study a spatial panel data model with fixed effects which incorporates spatial autoregression and spatial error structures simultaneously with possible different spatial weights matrices for these two components. They show that the MLE estimator of variance parameter (σ^2) is inconsistent if N is large and T is small. However, if the model also includes time effects, all parameters are inconsistently estimated using MLE even when both N and T are large. Instead, they propose a data transformation approach and establish consistency and asymptotic normality of the quasi-maximum likelihood estimators (QMLE) for either large N or large T.

All the aforementioned models are static panel data models. There has been increased interests in estimation of spatial dynamic panel data models over the past decade. Yu et al.

[1]In both SEM and SAR models, stationarity requires that $1/\lambda_{min}(W) < \rho < 1/\lambda_{max}(W)$ and $1/\lambda_{min}(W) < \delta < 1/\lambda_{max}(W)$, where, similar to the definition of $\lambda_{max}(A)$, $\lambda_{min}(A)$ denotes the minimum eigenvalue of A.

(2008) consider a dynamic SAR panel data model with fixed effects which includes individual time lags (that is, $\mathbf{y}_{t-1}$) and spatial time lags (that is, $W\mathbf{y}_{t-1}$). They investigate the asymptotic properties of QMLE for their model when both N and T are large. Moreover, Lee and Yu (2010b, 2014) extend this study to include time effects, where, in Lee and Yu (2014), a generalized method of moments (GMM) is developed. Su and Yang (2015) propose QMLE for a dynamic SEM panel data model when N is large and T is fixed. Both the random effects and fixed effects are considered in their model. The reader is referred to the paper by Lee and Yu (2010c) for more details about recent developments in spatial panel data models.

With regard to robust inference for spatial panel data models, there is a rich literature on estimation of covariance matrix which accounts for spatial correlation. In particular, recently, Bester et al. (2016) extend the time series fixed-b approach of Kiefer and Vogelsang (2005) to allow for spatial dependence.

3.3 Factor approaches

The spatial correlation is not the only pattern of cross-sectional dependence. For example, in macroeconomics, common shocks (e.g., financial crises, oil price shocks, technological shocks) may affect all cross-sectional units and thus cause cross-correlation. Factor approach, which assumes that the error terms contain a finite number of unobserved common factors that influence each individual, is widely used to model cross-sectional dependence of this kind. When N is large and T is small, Ahn et al. (2001) consider a linear panel data model with time-varying individual effects which are, in essence, a factor structure. They propose a number of GMM estimators that utilize the first- and second-order moment conditions implied by exogeneity of the regressors and by homoskedasticity and nonautocorrelation of the error terms. Bai (2003) develops an inferential theory for pure factor models of large N and T based on the principal components analysis (PCA) estimator. The recent literature mainly focuses on the estimation of the slope coefficients of regressors in linear panel data models with multifactor error structure when both N and T are large. For example, Pesaran (2006) provides the common correlated effect estimator (CCE) and Bai (2009) proposes the interactive fixed effects estimator (IFE), respectively.

3.3.1 CCE estimation

Pesaran (2006) considers the following heterogeneous panel data model

$$y_{it} = \alpha_i'\mathbf{d}_t + \beta_i'\mathbf{x}_{it} + e_{it} \quad \text{with} \quad e_{it} = \gamma_i'\mathbf{f}_t + \varepsilon_{it}, \tag{8}$$

where $\mathbf{d}_t$ is an $m \times 1$ vector of observed common factors which may include intercepts or seasonal dummies, and $\mathbf{f}_t$ is an $r \times 1$ vector of unobserved common factors. Chudik et al. (2011) show that the process $\{e_{it}\}$ is strongly dependent across i.

In general, regressors and unknown factors can be correlated. Simply regressing y_{it} on $\mathbf{x}_{it}$ leads to inconsistent estimation. The rationale of the CCE estimation is to replace (approximate) the unobserved factors $\mathbf{f}_t$ with observed variables. To this end, let

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \mathbf{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{9}$$

where $\mathbf{A}_i$ and $\mathbf{\Gamma}_i$ are $m \times p$ and $r \times p$ factor loadings of $\mathbf{x}_{it}$. Combining (8) and (9) yields

$$\begin{aligned}\mathbf{z}_{it} &= \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \begin{pmatrix} \alpha_i' + \beta_i'\mathbf{A}_i' \\ \mathbf{A}_i' \end{pmatrix} \mathbf{d}_t + \begin{pmatrix} \gamma_i' + \beta_i'\mathbf{\Gamma}_i' \\ \mathbf{\Gamma}_i' \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} \varepsilon_{it} + \beta_i'\mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix} \\ &\equiv \mathbf{B}_i'\mathbf{d}_t + \mathbf{C}_i'\mathbf{f}_t + \mathbf{u}_{it}. \end{aligned} \tag{10}$$

The weighted cross-sectional average of (10) leads to

$$\bar{\mathbf{z}}_{wt} = \bar{\mathbf{B}}_w'\mathbf{d}_t + \bar{\mathbf{C}}_w'\mathbf{f}_t + \bar{\mathbf{u}}_{wt},$$

where $\bar{\mathbf{z}}_{wt} = \sum_{i=1}^N w_i \mathbf{z}_{it}$ with some weights $\{w_i\}_{i=1}^N$ and $\bar{\mathbf{B}}_w,\ \bar{\mathbf{C}}_w,\ \bar{\mathbf{u}}_{wt}$ are defined accordingly. Therefore, if

$$\text{rank}(\bar{\mathbf{C}}_w) = r \le p+1, \tag{11}$$

we have, for each t,

$$\mathbf{f}_t - (\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}(\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}_w'\mathbf{d}_t) \xrightarrow{p} \mathbf{0}, \quad \text{as } N \to \infty \tag{12}$$

by the law of large numbers under certain regularity conditions, where $\mathbf{C} \equiv \operatorname*{plim}_{N\to\infty} \bar{\mathbf{C}}_w$. (12) implies that $\mathbf{f}_t$ can be approximated by a linear combination of $\bar{\mathbf{z}}_{wt}$ and $\mathbf{d}_t$. Thus, β_i can be consistently estimated by considering the following working model

$$y_{it} = \beta_i'\mathbf{x}_{it} + \vartheta_i'\mathbf{h}_{wt} + \varepsilon_{it}^*, \tag{13}$$

where $\mathbf{h}_{wt} = (\mathbf{d}_t', \bar{\mathbf{z}}_{wt}')'$. The CCE estimator of the individual slope coefficients is given by

$$\hat{\beta}_i = (\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{y}_i, \tag{14}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})'$, $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$, $\bar{\mathbf{M}}_w = I_T - \bar{\mathbf{H}}_w(\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w)^{-1}\bar{\mathbf{H}}_w'$ and $\bar{\mathbf{H}}_w = (\mathbf{h}_{w1}, \dots, \mathbf{h}_{wT})'$. If $\beta_i = \beta$ for all i, efficiency can be achieved by pooling all observations over the cross-sectional units. Such a pooled CCE (CCEP) estimator is given by

$$\hat{\beta}_{\text{CCEP}} = \left[\sum_{i=1}^N w_i \mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right]^{-1} \sum_{i=1}^N w_i \mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{y}_i. \tag{15}$$

Pesaran (2006) shows that, under some general conditions, $\hat{\beta}_i$ and $\hat{\beta}_{\text{CCEP}}$ are $\sqrt{T}$ and $\sqrt{NT}$ consistent, respectively, and both estimators are asymptotically normally distributed, provided that the rank condition in (11) is satisfied; see Theorems 1 and 4 in Pesaran(2006).

3.3.2 IFE estimation

When $\beta_i = \beta$ and $\alpha_i = \mathbf{0}$ for all i, the model setup in (8) corresponds to the model in Bai (2009), where the term "interactive fixed effects" is employed to indicate that $\mathbf{f}_t$ and γ_i enter into the model multiplicatively. Instead of filtering unobserved common factors by cross-sectional average of observed variables as in Pesaran (2006), Bai (2009) proposes to estimating the factors and factor loadings together with the structural parameter β using PCA, which requires a priori knowledge of the number of factors. Specifically, let $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)'$ and

$\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_N)'$, which satisfy

$$\mathbf{F}'\mathbf{F}/T = I_r, \quad \text{and} \quad \mathbf{\Gamma}'\mathbf{\Gamma} = \text{diagonal}.$$

Define

$$M_{\mathbf{F}} = I_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}' = I_T - \mathbf{F}\mathbf{F}'/T.$$

Then, the estimator $(\hat{\beta}, \hat{\mathbf{F}})$ is the solution of the set of nonlinear equations

$$\hat{\beta} = (\sum_{i=1}^{N} \mathbf{X}_i' M_{\hat{\mathbf{F}}} \mathbf{X}_i)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' M_{\hat{\mathbf{F}}} \mathbf{y}_i$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{X}_i\hat{\beta})(\mathbf{y}_i - \mathbf{X}_i\hat{\beta})'\right]\hat{\mathbf{F}} = \hat{\mathbf{F}} V_{NT},$$

where V_{NT} is a diagonal matrix which consists of the r largest eigenvalues of $\frac{1}{NT}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{X}_i\hat{\beta})(\mathbf{y}_i - \mathbf{X}_i\hat{\beta})'$ arranged in decreasing order. Therefore, given $\mathbf{F}$, one can estimate β, and given β, one can estimate $\mathbf{F}$. The solution $(\hat{\beta}, \hat{\mathbf{F}})$ can be simply obtained by an iterative algorithm.

Bai (2009) shows that, under some regularity conditions, $\hat{\beta}$ is $\sqrt{NT}$ consistent and is asymptotically normally distributed. However, the limiting distribution will not be centered at zero when correlation and heteroskedasticity of ε_{it} present in both dimensions. Furthermore, the zero mean asymptotic distribution can be achieved when there is no heteroskedasticity and correlation is absent in at least one dimension; see Theorems 2 and 3 in Bai (2009) for details.

3.3.3 Further developments

The CCE approach has gained a lot of attentions during the last a few years, which does not require a priori knowledge of the number of unobserved common factors and leads to a simple OLS estimation of an augmented regression. A number of papers has extended the CCE type estimators in several dimensions. For example, Pesaran and Tosetti (2011) show the consistency and asymptotic normality for the CCE estimators of the slope coefficients when $\{\varepsilon_{it}\}$ in (8) are generated by a spatial process. Kapetanios et al. (2011) extend the analysis of Pesaran (2006) to the case where the unobserved common factors are integrated of order 1. By distinguishing among the concepts of weak, strong and semi-strong common factors based on Definition 2.1, Chudik et al. (2011) study the CCE estimation of slope coefficients when the errors consist of a finite number of strong factors and an infinite number of weak and/or semi-strong factors. Westerlund and Urbain (2013) show the inconsistency of the CCEP estimator when factor loadings γ_i in (8) and $\mathbf{\Gamma}_i$ in (9) are correlated. Finally, Harding and Lamarche (2011) introduce endogeneity into model (8) and (9) by allowing ε_{it} and $\mathbf{v}_{it}$ to be correlated, and develop a two-step instrumental variables CCE estimation procedure for the homogeneous slope coefficients.

In some empirical studies, researchers may also be interested in the estimation of latent factors and their loadings. However, both of them are treated as nuisance parameters in the CCE approach. In order to extract the information about factor structures, a two-step procedure is proposed by Castagnetti et al. (2015), in which the first step is to estimate the slope coefficients using the CCE estimator, and then the second step is to compute the residuals to which the PCA is applied to obtain the estimators of factors and loadings. Greenaway-McGrevy et al. (2012), instead, try to uncover the conditions under which factor estimates using PCA can be used to replace the common factors without affecting the limiting distribution of the slope coefficients. They find that $T/N \to 0$ and $N/T^3 \to 0$ are sufficient for this replacement under some regularity conditions.

Also, Everaert and Groote (2016) extend the analysis of Pesaran (2006) by investigating the asymptotic properties of the CCEP estimator in a dynamic panel data setting. They point out that the CCEP estimator is no longer consistent for $N \to \infty$ and fixed T, and derive the asymptotic bias of the CCEP estimator in the case of dynamic homogeneous panel data models. Chudik and Pesaran (2015) propose a dynamic CCE estimation approach for the following heterogeneous dynamic panel data model with weakly exogenous regressors

$$\begin{aligned} y_{it} &= \alpha_{yi} + \phi_i y_{i,t-1} + \beta_{0i}' \mathbf{x}_{it} + \beta_{1i}' \mathbf{x}_{i,t-1} + e_{it} \quad \text{with} \quad e_{it} = \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \\ \omega_{it} &= \begin{pmatrix} \mathbf{x}_{it} \\ \mathbf{g}_{it} \end{pmatrix} = \alpha_{\omega i} + \vartheta_i y_{i,t-1} + \mathbf{\Gamma}_i' \mathbf{f}_t + \mathbf{v}_{it}, \end{aligned}$$

where α_{yi} and $\alpha_{\omega i}$ are individual-specific fixed effects, $\mathbf{x}_{it}$ is $p_x \times 1$ vector of regressors, $\mathbf{g}_{it}$ is $p_g \times 1$ vector of additional covariates that are affected by the same set of unobserved common factors, ϑ_i is a $(p_x + p_g) \times 1$ vector of unknown feed-back coefficients that can be used to distinguish between strictly ($\vartheta_i = \mathbf{0}$) and weakly exogenous regressors. Performing similar procedure as in Section 3.3.1, Chudik and Pesaran (2015) show that under some general conditions,

$$y_{it} = \alpha_{yi}^* + \phi_i y_{i,t-1} + \beta_{0i}' \mathbf{x}_{it} + \beta_{1i}' \mathbf{x}_{i,t-1} + \delta_i'(L) \bar{\mathbf{z}}_{wt} + \varepsilon_{it} + O_p(N^{-1/2}),$$

where $\delta_i(L) = \sum_{l=0}^{\infty} \delta_{il} L^l$, $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}_{wt}', \bar{\mathbf{g}}_{wt}')' = \sum_{i=1}^{N} w_i \mathbf{z}_{it}$ and $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}', \mathbf{g}_{it}')'$, provided that the number of cross-sectional averages is at least as large as the number of unobserved common factors. This result indicates that a sufficient number of lags of cross-sectional averages $\bar{\mathbf{z}}_{wt}$ must be included in the augmented regression. Therefore, the dynamic CCE estimator of $\pi_i \equiv (\phi_i, \beta_{0i}', \beta_{1i}')'$ is obtained by considering the OLS estimation of the following augmented regression

$$y_{it} = \alpha_{yi}^* + \phi_i y_{i,t-1} + \beta_{0i}' \mathbf{x}_{it} + \beta_{1i}' \mathbf{x}_{i,t-1} + \sum_{l=0}^{k_T} \delta_{il}' \bar{\mathbf{z}}_{w,t-l} + \varepsilon_{it}^*,$$

where k_T is the number of lags. Define

$$\mathbf{\Xi}_i = \begin{pmatrix} y_{ik_T} & \mathbf{x}_{i,k_T+1}' & \mathbf{x}_{i,k_T}' \\ y_{i,k_T+1} & \mathbf{x}_{i,k_T+2}' & \mathbf{x}_{i,k_T+1}' \\ \vdots & \vdots & \vdots \\ y_{i,T-1} & \mathbf{x}_{iT}' & \mathbf{x}_{i,T-1}' \end{pmatrix}, \quad \bar{\mathbf{Q}}_w = \begin{pmatrix} 1 & \bar{\mathbf{z}}_{w,k_T+1}' & \bar{\mathbf{z}}_{w,k_T}' & \cdots & \bar{\mathbf{z}}_{w,1}' \\ 1 & \bar{\mathbf{z}}_{w,k_T+2}' & \bar{\mathbf{z}}_{w,k_T+1}' & \cdots & \bar{\mathbf{z}}_{w,2}' \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \bar{\mathbf{z}}_{w,T}' & \bar{\mathbf{z}}_{w,T-1}' & \cdots & \bar{\mathbf{z}}_{w,T-k_T}' \end{pmatrix},$$

and

$$\bar{\mathbf{M}}_q = I_{T-k_T} - \bar{\mathbf{Q}}_w (\bar{\mathbf{Q}}_w' \bar{\mathbf{Q}}_w)^+ \bar{\mathbf{Q}}_w'$$

where "+" denotes the Moore-Penrose generalized inverse. A routine computation gives rise to

$$\hat{\pi}_i = (\mathbf{\Xi}_i' \bar{\mathbf{M}}_q \mathbf{\Xi}_i)^{-1} \mathbf{\Xi}_i' \bar{\mathbf{M}}_q \mathbf{y}_i,$$

where $\mathbf{y}_i = (y_{i,k_T+1}, y_{i,k_T+2}, \ldots, y_{i,T})'$. The estimator of $\pi(\equiv E(\pi_i))$ is given by

$$\hat{\pi}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\pi}_i,$$

which is called the mean group (MG) estimator. Chudik and Pesaran (2015) show that $\hat{\pi}_i$ and $\hat{\pi}_{MG}$ are consistent estimators of π_i and π, respectively, provided that the $(p_x+p_g+1)\times r$ matrix $\mathbf{C} = (E(\gamma_i), E(\mathbf{\Gamma}_i))'$ is full of column rank and $k_T^3/T \to \chi,\ 0 < \chi < \infty$ as $(N, T, k_T) \to \infty$. If the rank condition does not hold, $\hat{\pi}_i$ is inconsistent, and $\hat{\pi}_{MG}$ is still consistent, as long as $\mathbf{f}_t$ is serially uncorrelated. Chudik and Pesaran (2015) further prove that $\hat{\pi}_{MG}$ is asymptotic normality as $(N, T, k_T) \to \infty$ such that $N/T \to \chi_1$ and $k_T^3/T \to \chi_2$, $0 < \chi_1, \chi_2 < \infty$. However, the convergence rate of $\hat{\pi}_{MG}$ is $\sqrt{N}$.

Another practical issue in implementing the CCE approach is related to the rank condition in (11). Although no information about the number of factors is required when applying the CCE approach, (11) implies that choosing the number of regressors (p) actually makes restriction on the number of common factors (r). Indeed, Pesaran (2006) assumes that the slope coefficients β_i follow a random coefficient specification

$$\beta_i = \beta + \upsilon_i, \quad \upsilon_i \sim \text{IID}(\mathbf{0}, \mathbf{\Omega}_\upsilon),$$

for $i = 1, 2, \ldots, N$, where $\mathbf{\Omega}_\upsilon$ is a symmetric nonnegative definite matrix. The cross-sectional mean of β_i, namely β, can be estimated by the MG estimator; that is,

$$\hat{\beta}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i,$$

where $\hat{\beta}_i$ is defined in (14). Pesaran (2006) proves that when the rank condition (11) is not satisfied, $\hat{\beta}_i$ is inconsistent. However, $\hat{\beta}_{MG}$ is still $\sqrt{N}$-consistent and asymptotically normally distributed regardless of (11), as long as γ_i and $\mathbf{\Gamma}_i$ follow similar random coefficient models as β_i. Similar result holds for the CCEP estimator if it is used to estimate the expectation of β_i. Unfortunately, if β_i's are homogeneous ($\mathbf{\Omega}_\upsilon = \mathbf{0}$), there are no general results available for the CCEP estimator except for the special case when $r = 1$; see Theorem 4 in Pesaran (2006). Indeed, as what Karabiyik et al. (2014) point out, it is not clear from Pesaran (2006) whether one can permit $p+1 < r$ when $\mathbf{\Omega}_\upsilon = \mathbf{0}$ in order to achieve a faster convergence rate. Karabiyik et al. (2014) show that, if $p + 1 < r$, the rate of consistency is at best $\sqrt{N}$. Furthermore, if γ_i is non-IID and/or correlated with $\mathbf{\Gamma}_i$, then the CCEP estimator for the homogeneous slope coefficients is inconsistent. Therefore, Karabiyik et al. (2014) propose a combination-augmented CCE, termed as C^3E to overcome the limitation of (11). The idea behind this method is to augment the regression in (13) with different combinations of $\{\mathbf{z}_{1t}, \ldots, \mathbf{z}_{Nt}\}$, where $\mathbf{z}_{it}$ is defined in (10).

On the other hand, there have also been some papers to extend the work in Bai (2009) to many dimensions. Recall that Bai (2009) assumes that the error term ε_{it} is independent of regressors $\mathbf{x}_{js}$, factor loadings γ_j and common factors $\mathbf{f}_s$ for all i, t, j and s, which rules out the possibility of dynamic panel data models, and the number of factors is assumed to be known. To relax these assumptions, Moon and Weidner (2013) relax the strict exogeneity assumption by allowing some of the regressors to be predetermined but still assuming that the number of factors is known. They propose the Gaussian QMLE of the homogeneous slope coefficients, denoted as $\hat{\beta}_{QMLE}$. Moon and Weidner (2013) show that $\hat{\beta}_{QMLE}$ is consistent as $(N,T)\to\infty$ without any restrictions on the ratio N/T. The asymptotic distribution of $\hat{\beta}_{QMLE}$ is derived under additional requirements which include $N/T\to\kappa^2,\ 0<\kappa<\infty$ and cross-sectional independence of $\{(\mathbf{x}_{it},\varepsilon_{it}),t=1,2,\dots,T\}$ conditional on the σ-algebra generated by $\{(\gamma_i,\mathbf{f}_t),i=1,2,\dots,N;t=1,2,\dots,T\}$. It turns out that $\hat{\beta}_{QMLE}$ is asymptotically biased and there are two sources of the bias. The first is the correlation or heteroskedasticity of the error terms $\{\varepsilon_{it}\}$, which has already manifested itself in Bai (2009), and the second is due to the predetermined regressors. Furthermore, Moon and Weidner (2015) relax the requirement of knowing the number of factors as priori. They investigate the asymptotic properties of the QMLE when the true number of factors r is unknown and $r^*(\geq r)$ number of factors are used in the estimation. Let this estimator denoted by $\hat{\beta}^*_{QMLE}$. Under some conditions such as the error terms are IID and follow normal distribution, they derive the asymptotic normality of the proposed estimator as follows:

$$\sqrt{NT}(\hat{\beta}^*_{QMLE}-\beta)=\sqrt{NT}(\hat{\beta}_{QMLE}-\beta)+o_p(1)$$

as $(N,T)\to\infty$ with $N/T\to\kappa^2,\ 0<\kappa<\infty$, which implies that $\hat{\beta}^*_{QMLE}$ has the same limiting distribution as $\hat{\beta}_{QMLE}$.

Song (2013) extends the IFE approach in Bai (2009) to the case of heterogeneous dynamic panel data models with interactive fixed effects. The estimator $(\{\hat{\beta}_i\}_{i=1}^N,\hat{\mathbf{F}})$ can be obtained by solving the following set of nonlinear equations

$$\hat{\beta}_i=(\mathbf{X}_i'M_{\hat{\mathbf{F}}}\mathbf{X}_i)^{-1}\mathbf{X}_i'M_{\hat{\mathbf{F}}}\mathbf{y}_i$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^N(\mathbf{y}_i-\mathbf{X}_i\hat{\beta}_i)(\mathbf{y}_i-\mathbf{X}_i\hat{\beta}_i)'\right]\hat{\mathbf{F}}=\hat{\mathbf{F}}V_{NT},$$

where V_{NT} is a diagonal matrix which consists of the r largest eigenvalues of $\frac{1}{NT}\sum_{i=1}^N(\mathbf{y}_i-\mathbf{X}_i\hat{\beta}_i)(\mathbf{y}_i-\mathbf{X}_i\hat{\beta}_i)'$ arranged in decreasing order. Song (2013) establishes the consistency of $\hat{\beta}_i$ as $(N,T)\to\infty$. The asymptotic normality with convergence rate $\sqrt{T}$ is also derived under some additional assumptions including independence of $\{\varepsilon_{it},t=1,2,\dots,T\}$ over i and $T/N^2\to0$.

From the above discussions of two major estimations: CCE and IFE, a natural question that arises is, for a given application, which method should be used in practice. To answer this question, it is interesting to note that, Westerlund and Urbain (2015) recently provide a formal comparison between CCE and IFE by considering "the same data generating process and the same implementation approach, but different factor estimates", which implies that the estimators they consider are not identical to those considered in Pesaran (2006) and Bai (2009). Therefore, they actually compare the relative merits of two methods of estimating the unobserved common factors: cross-sectional averages and principal components. As a result, they remind researchers not to extrapolate too widely the conclusions drawn from their paper. In sum, they find that, if $T/N\to0$, these two estimators are asymptotically equivalent. However, when $T/N\to\chi>0$, the performance of the two estimators relies on the value of β.

IFE is subject to a relatively small bias when $\beta = \mathbf{0}$. Otherwise, CCE is expected to outperform IFE.

§4 Nonparametric and semiparametric panel data models with cross-sectional dependence

It is clear that all of the aforementioned methods focus on the linear specification of regression relationship. However, these simple parametric panel data models may be misspecified, which possibly results in misleading inference. There exists a rich literature on nonparametric and semiparametric panel data models which assume independence among cross-sectional units. To name just a few examples, Cai and Li (2008), Henderson et al. (2008) and Lin et al. (2014) and the references therein.

The first nonparametric panel data model with cross-sectional dependence, to the best of our knowledge, is the following model proposed by Su and Jin (2012), given by

$$y_{it} = g_i(\mathbf{x}_{it}) + \alpha_i'\mathbf{d}_t + e_{it} \quad \text{with} \quad e_{it} = \gamma_i'\mathbf{f}_t + \varepsilon_{it}, \tag{16}$$

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \mathbf{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{17}$$

where $g_i(\cdot)$ is an unknown continuous function from $\mathbb{R}^p$ to $\mathbb{R}$, which may assume different form for different i. It is apparent from (8) and (9) that the above model includes the models in Pesaran (2006) and Bai (2009) as a special case. Su and Jin (2012) propose using a sieve estimation for the nonparametric heterogeneous functions by extending the CCE approach to this nonparametric framework. They find that $\mathbf{h}_t = (\mathbf{d}_t', \bar{\mathbf{z}}_t')'$ can still be used as observable proxies for $\mathbf{f}_t$ under regularity conditions, where $\bar{\mathbf{z}}_t = \frac{1}{N}\sum_{i=1}^N \mathbf{z}_{it}$ given in (10). Let $\{p_l(\mathbf{x}), l = 1, 2, \ldots\}$ denote a sequence of known basis functions employed in sieve estimation. Therefore, for any given grid point $\mathbf{x}$, $g_i(\mathbf{x})$ is approximated by $\alpha_{g_i}' p^K(\mathbf{x})$, where $p^K(\mathbf{x}) = (p_1(\mathbf{x}), \ldots, p_K(\mathbf{x}))$ and $K \to \infty$ as $T \to \infty$. To estimate α_{g_i}, the following augmented regression is considered

$$y_{it} = \alpha_{g_i}' p^K(\mathbf{x}_{it}) + \vartheta_i'\mathbf{h}_t + \varepsilon_{it}^*. \tag{18}$$

By rewriting (18) in a vector form, we obtain

$$\mathbf{y}_i = p_i\alpha_{g_i} + \mathbf{H}\vartheta_i + \varepsilon_i^*,$$

where $p_i = (p_{i1}, \ldots, p_{iT})'$, $p_{it} = p^K(\mathbf{x}_{it})$ and $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_T)'$. It follows that

$$\hat{\alpha}_{g_i} = (p_i'\mathbf{M}p_i)^+ p_i'\mathbf{M}\mathbf{y}_i$$

and

$$\hat{g}_i(\mathbf{x}) = p^K(\mathbf{x})'\hat{\alpha}_{g_i},$$

where $\mathbf{M} = I_T - \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$. If $g_i(\cdot) = g(\cdot)$ for all i, a similar argument leads to

$$\hat{\alpha}_g = \left(\sum_{i=1}^N p_i'\mathbf{M}p_i\right)^+ \sum_{i=1}^N p_i'\mathbf{M}\mathbf{y}_i$$

so that $\hat{g}(\mathbf{x}) = p^K(\mathbf{x})'\hat{\alpha}_g$. Su and Jin (2012) derive the convergence rate and the asymptotic normality for both $\hat{g}_i(\mathbf{x})$ and $\hat{g}(\mathbf{x})$ when both N and T are large. It is well known that a sieve method is a global approximation which can not well capture the local properties of functionals.

Huang (2013) also considers the model in (16) with $g_i(\cdot) = g(\cdot)$ for all i, without imposing (17) and mainly concentrates on the case where $\mathbf{d}_t = 1$ under which the model reduces to

$$y_{it} = g(\mathbf{x}_{it}) + \alpha_i + \gamma_i'\mathbf{f}_t + \varepsilon_{it}, \tag{19}$$

where, if correlated with $\mathbf{x}_{it}$, α_i is treated as fixed effect. Following a similar argument as in Pesaran (2006), Huang (2013) shows that the cross-sectional averages $\bar{y}_t = \frac{1}{N}\sum_{i=1}^N y_{it}$ and $\bar{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_{it}$ can be used to filter unobserved common factors and proposes to estimating $g(\mathbf{x}_{it})$ using a local linear approach to capture the local properties of functionals. Without assuming $\frac{1}{N}\sum_{i=1}^N \alpha_i = 0$, $g(\cdot)$ is not identified. He further shows that $\bar{y}_i = \frac{1}{T}\sum_{t=1}^T y_{it}$ and $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^T \mathbf{x}_{it}$ can be used to filter fixed effects. Based on these facts, Huang (2013) indeed considers the following oversimplified model with common parameters across i; that is,

$$y_{it} = g(\mathbf{x}_{it}) + \beta_1\bar{y}_t + \beta_2'\bar{\mathbf{x}}_t + \beta_3\bar{y}_i + \beta_4'\bar{\mathbf{x}}_i + \varepsilon_{it}^*,$$

and derives the consistency and the asymptotic normality of the local linear estimator of $g(\cdot)$ whose convergence rate relies on NT and bandwidth h.

Su and Zhang (2013) consider homogeneous model (19) with $\alpha_i = 0$ and propose using sieve estimation approach together with the QMLE in Moon and Weidner (2013) to estimate the nonparametric function $g(\cdot)$. They derive the convergence rate for the sieve based QMLE estimator, $\tilde{g}(\cdot)$, under some general conditions which are widely used in the literature on panel data models with factor structure. To establish the asymptotic normality of $\tilde{g}(\cdot)$, they further impose conditionally cross-sectional independence between $(\mathbf{x}_{it}, \varepsilon_{it})$ and $(\mathbf{x}_{js}, \varepsilon_{js})$ for all $i \neq j$ and all $t, s = 1, 2, \ldots, T$, and conditionally strong mixing of $\{(\mathbf{x}_{it}, \varepsilon_{it}), t = 1, 2, \ldots, T\}$ for each i, conditional on the σ-algebra generated by $\{(\gamma_i, \mathbf{f}_t), i = 1, 2, \ldots, N; t = 1, 2, \ldots, T\}$. The asymptotic distribution of $\tilde{g}(\cdot)$ involves a bias term based on what they propose a bias-corrected estimator for $g(\cdot)$, which is denoted as $\tilde{g}_{bc}(\cdot)$. They also propose a specification test for the linearity of the functional form by considering the following null hypothesis and sequences of Pitman local alternatives:

$$\mathbb{H}_0: \ \Pr\left[g(\mathbf{x}_{it}) = \mathbf{x}_{it}'\beta_0\right] = 1 \text{ for some } \beta_0 \in \mathbb{R}^p \quad \text{versus} \quad \mathbb{H}_1: g(\mathbf{x}_{it}) = \mathbf{x}_{it}'\beta_0 + \gamma_{NT}\Delta(\mathbf{x}_{it}),$$

where $\Delta(\cdot)$ is a measurable nonlinear function and $\gamma_{NT} \to 0$ as $(N, T) \to \infty$. They construct the test statistic based on the L_2-distance between the linear and bias-corrected estimators, i.e.,

$$\Gamma_{NT} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left[\tilde{g}_{bc}(\mathbf{x}_{it}) - \mathbf{x}_{it}'\hat{\beta}_{QMLE}\right]^2 w(\mathbf{x}_{it}),$$

where $w(\cdot)$ is a nonnegative weighting function. After being appropriately normalized, they show that Γ_{NT} follows the standard normal distribution asymptotically and has nontrivial power to detect sequences of Pitman local alternatives that converge to the null at certain rate. Su et al. (2015) also propose a residual-based test for $\mathbb{H}_0$ against $\mathbb{H}_1$ in panel data models with interactive fixed effects.

Models given in (16) and (19) are fully nonparametric with respect to $g_i(\cdot)$ and $g(\cdot)$, respectively. It is well known that nonparametric estimation suffers the so-called "curse of dimensionality" and lacks economic explanations. To overcome this difficulty, some dimension reduction models are developed recently. For example, recently, Dong et al. (2015) consider a semiparametric single-index panel data model which is specified as follows:

$$y_{it} = g(\mathbf{x}_{it}'\theta_0) + \alpha_i + e_{it},$$

where $g(\cdot)$ is an unknown link function. The cross-sectional dependence among $\{\mathbf{x}_{it}, e_{it}\}$ is characterized using a general *spatial mixing* structure that integrates the correlation across individuals and time. Let $\mathbf{x}_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{Nt})'$ and $e_t = (e_{1t}, \ldots, e_{Nt})'$. The process $\{(\mathbf{x}_t, e_t) : 1 \le t \le T\}$ is assumed to be strictly stationary and α-mixing. Dong et al. (2015) impose the following conditions on the α-mixing coefficient $\alpha_{ij}(|t-s|)$ between $\{\mathbf{x}_{it}, e_{it}\}$ and $\{\mathbf{x}_{js}, e_{js}\}$:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{\infty}(\alpha_{ij}(t))^{\eta/(4+\eta)} = O(N), \quad \text{and} \quad \sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_{ij}(0))^{\eta/(4+\eta)} = O(N), \tag{20}$$

where $\eta > 0$ is chosen such that $E\left[|e_{it}|^{4+\eta}\right] < \infty$ and $E\left[\|\mathbf{x}_{it}\|^{4+\eta}\right] < \infty$. It is to easy to see that the strength of correlation among both dimensions is controlled by the first equation in (20), while the strength of cross-sectional dependence at any given time is controlled by the second equation in (20). Dong et al. (2015) use a single factor model structure to show that the above conditions are verifiable. However, this requires that γ_i converges to 0 at a certain rate as i increases.

Finally, Cai et al. (2016) consider a varying-coefficient panel data model that $g_i(\mathbf{x}_{it})$ is replaced by $\beta_i(\mathbf{u}_{it})'\mathbf{x}_{it}$ in (16), where $\mathbf{u}_{it} \in \mathbb{R}^d$ is a vector of smooth variables. To allow for the possibility of correlation between $\mathbf{u}_{it}$ and common factors, Cai et al. (2016) adopt the following fairly general model for both $\mathbf{x}_{it}$ and $\mathbf{u}_{it}$,

$$\mathbf{z}_{it} = \begin{pmatrix} \mathbf{x}_{it} \\ \mathbf{u}_{it} \end{pmatrix} = \mathbf{A}_i'\mathbf{d}_t + \mathbf{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}.$$

Based on the idea of CCE approach, they show that $\mathbf{d}_t$ and the cross-sectional average of $\mathbf{z}_{it}$ can be utilized as observable proxies for $\mathbf{f}_t$, provided that

$$\text{rank}[E(\mathbf{\Gamma}_i)] \le p + d.$$

They study the estimation of both heterogeneous and homogeneous coefficient functions using local linear estimation approach. For the estimation of $\beta_i(\cdot)$, they consider the following augmented regression

$$y_{it} = \beta_i'(\mathbf{u}_{it})\mathbf{x}_{it} + \vartheta_i'\mathbf{q}_t + \varepsilon_{it}^*,$$

where $\mathbf{q}_t = (\mathbf{d}_t', \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_{it}')'$. Then, for any fixed $u_0 \in \mathbb{R}$, $\beta_i(u_0)$ and its first order derivative $\beta_i^{(1)}(u_0)$ can be estimated by [2]

$$\begin{aligned}\hat{\beta}_i^*(u_0) &\equiv \begin{pmatrix} \hat{\beta}_i(u_0) \\ h\hat{\beta}_i^{(1)}(u_0) \end{pmatrix} \\ &= \arg\min_{a,b}\sum_{t=1}^{T}\left[y_{it} - (\mathbf{x}_{it}', \mathbf{x}_{it}'(\frac{\mathbf{u}_{it}-u_0}{h}))\begin{pmatrix} a \\ b \end{pmatrix} - \vartheta_i'\mathbf{q}_t\right]^2 k_h(\mathbf{u}_{it}-u_0),\end{aligned}$$

[2] For ease of notation, Cai et al. (2016) only consider the case $d = 1$. They point out that extension to the case $d > 1$ involves no fundamentally new ideas and models with large d are not practically useful due to "curse of dimensionality".

where h is the bandwidth, $k(\cdot)$ is a kernel function and $k_h(\cdot) = k(\cdot/h)/h$. Note that the above local linear estimator can be viewed as the OLS estimator of the working linear model

$$\mathbf{K}_{i,h}^{1/2}(u_0)\mathbf{y}_i = \mathbf{K}_{i,h}^{1/2}(u_0)\widetilde{\mathbf{X}}_i\beta_i^*(u_0) + \mathbf{K}_{i,h}^{1/2}(u_0)\mathbf{Q}\vartheta_i + \mathbf{K}_{i,h}^{1/2}(u_0)\varepsilon_i^*,$$

where $\mathbf{K}_{i,h}(u_0) = diag(k_h(\mathbf{u}_{i1} - u_0), \ldots, k_h(\mathbf{u}_{iT} - u_0))$, $\widetilde{\mathbf{X}}_i$ is the $T \times 2p$ matrix whose t-th row is $(\mathbf{x}_{it}', \mathbf{x}_{it}'(\frac{\mathbf{u}_{it}-u_0}{h}))$, and $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_T)'$. Let

$$M_i(u_0) = I_T - \mathbf{K}_{i,h}^{1/2}(u_0)\mathbf{Q}[\mathbf{Q}'\mathbf{K}_{i,h}(u_0)\mathbf{Q}]^{-1}\mathbf{Q}'\mathbf{K}_{i,h}^{1/2}(u_0).$$

By the formula for partitioned regression, one can easily obtain

$$\hat{\beta}_i^*(u_0) = \left[\widetilde{\mathbf{X}}_i'\mathbf{K}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{K}_{i,h}^{1/2}(u_0)\widetilde{\mathbf{X}}_i\right]^{-1}\widetilde{\mathbf{X}}_i'\mathbf{K}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{K}_{i,h}^{1/2}(u_0)\mathbf{y}_i.$$

By the same token, Cai et al. (2016) propose a CCEP type estimator for the homogeneous coefficient function $\beta(\cdot)$ and its derivative, which is given by

$$\begin{aligned}\hat{\beta}^*(u_0) &\equiv \begin{pmatrix} \hat{\beta}(u_0) \\ h\hat{\beta}^{(1)}(u_0) \end{pmatrix} \\ &= \left[\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\mathbf{K}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{K}_{i,h}^{1/2}(u_0)\widetilde{\mathbf{X}}_i\right]^{-1}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\mathbf{K}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{K}_{i,h}^{1/2}(u_0)\mathbf{y}_i.\end{aligned}$$

Cai et al. (2016) show that $\hat{\beta}_i(u_0)$ and $\hat{\beta}(u_0)$ are $\sqrt{Th}$ and $\sqrt{NTh}$ consistent, respectively, and both estimators are asymptotic normality when $(N, T) \to \infty$ at certain rate; see Theorems 4.2 and 5.2 in Cai et al. (2016).

More importantly, Cai et al. (2016) propose a novel nonparametric test for a parametric specification of the homogeneous coefficient function $\beta(\cdot)$ against a nonparametric alternative. Specifically, they consider the following null and alternative hypotheses

$$\mathbb{H}_0: \ \beta(u) = \beta_0(u) \quad \text{versus} \quad \mathbb{H}_1: \ \beta(u) \neq \beta_0(u),$$

where $\beta_0(u)$ is an known parametric function of u. They construct the test statistic based on the integrated squared difference between the parametric specification and varying coefficient; that is, $L \equiv \int[\beta(u) - \beta_0(u)]'[\beta(u) - \beta_0(u)]du$. A feasible test statistic can be obtained by replacing $\beta(u)$ with $\hat{\beta}(u)$ and replacing $\beta_0(u)$ with a $\sqrt{NT}$ consistent estimate $\hat{\beta}_0(u)$. After several steps of simplification, they obtain a test statistic which is of the following form

$$\hat{L}_{NT} = \frac{1}{N^2T^2h}\sum_{i=1}^N\sum_{j\neq i}^N \tilde{\nu}_i'\mathbf{K}(\mathbf{z}_i, \mathbf{z}_j)\tilde{\nu}_j,$$

where $\tilde{\nu}_i = (\tilde{\nu}_{i1}, \ldots, \tilde{\nu}_{iT})' = M_Q(\mathbf{y}_i - \mathcal{F}_i(\mathbf{X}_i, \mathbf{U}_i))$, $M_Q = I_T - \mathbf{Q}[\mathbf{Q}'\mathbf{Q}]^{-1}\mathbf{Q}'$, $\mathcal{F}_i(\mathbf{X}_i, \mathbf{U}_i) = (\mathbf{x}_{i1}'\hat{\beta}_0(u_{i1}), \ldots, \mathbf{x}_{iT}'\hat{\beta}_0(u_{iT}))'$, and $\mathbf{K}(\mathbf{z}_i, \mathbf{z}_j)$ is a matrix whose (t, s)th element is $\mathbf{x}_{it}'\mathbf{x}_{js}k(\frac{u_{it}-u_{js}}{h})$. Under certain regularity conditions, they show that $(NTh^{1/2})\hat{L}_{NT}/\sqrt{\hat{V}}$ is asymptotically standard normal under $\mathbb{H}_0$, where

$$\hat{V} = \frac{2}{N^2T^2h}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{1\leq t,s\leq T}\tilde{\nu}_{it}^2\tilde{\nu}_{js}^2(\mathbf{x}_{it}'\mathbf{x}_{js})^2k^2(\frac{u_{it}-u_{js}}{h})$$

is a consistent estimator of the asymptotic variance of $(NTh^{1/2})\hat{L}_{NT}$. Also, they establish the consistency of the test statistic under fixed alternatives.

In other directions, there are also a number of papers in the nonparametric and semiparametric literature to address cross-sectional dependence by directly imposing certain moment conditions on error terms. For example, Robinson (2012) was the first to consider the following model

$$y_{it} = \alpha_i + \beta_t + e_{it}, \tag{21}$$

where α_i and β_t are individual and time effects, and e_{it} are unobserved random variables such that $E(e_{it}) = 0$ for all i, t; $E(e_{it}e_{jt}) = \omega_{ij}$ for all i, j, t; $E(e_{it}e_{ju}) = 0, t \neq u$ for all i, j, t, u. The main interest in Robinson (2012) is to estimating the time trend, which is represented by β_t, when T is large and N is fixed. Taking cross-sectional average of (21) and imposing $\sum_{i=1}^N \alpha_i = 0$, one can get

$$\bar{y}_t = \beta_t + \bar{e}_t,$$

where $\bar{y}_t = N^{-1}\sum_{i=1}^N y_{it}$ and $\bar{e}_t = N^{-1}\sum_{i=1}^N e_{it}$. Robinson (2012) points out that $\bar{y}_t$ is a mean-square consistent estimator for β_t if N is large and the degree of cross-sectional dependence is limited by

$$\overline{\lim_{N\to\infty}} \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \omega_{ij} < \infty. \tag{22}$$

Unfortunately, (22) does not hold for the multifactor error structure model given in (8), unless further restrictions are imposed on the factor loadings. Instead of estimating β_t using $\bar{y}_t$, Robinson (2012) treats β_t as a smooth function of t; that is, $\beta_t = \beta(t/T)$, and estimates $\beta(\cdot)$ using a kernel method.

Chen et al. (2012) extend the work of Robinson (2012) to a semiparametric partially linear panel data model with cross-sectional dependence,

$$y_{it} = \mathbf{x}_{it}'\beta + f_t + \alpha_i + e_{it}, \quad \text{and} \quad \mathbf{x}_{it} = g_t + \chi_i + \mathbf{v}_{it},$$

where $f_t = f(t/T)$ and $g_t = g(t/T)$ are unknown time trend functions, α_i and χ_i are individual specific effects that satisfy $\sum_{i=1}^N \alpha_i = 0$ and $\sum_{i=1}^N \chi_i = \mathbf{0}$, respectively. A pooled semiparametric profile likelihood dummy variable estimation method is developed to estimate β and $f(\cdot)$. To this end, based on the local linear estimation approach, for a given grid point τ, $f(\tau)$ and its first order derivative $f^{(1)}(\tau)$ can be estimated by

$$\begin{pmatrix} \hat{f}_{\alpha,\beta}(\tau) \\ \hat{f}^{(1)}_{\alpha,\beta}(\tau) \end{pmatrix} = \arg\min_{(a,b)'} \sum_{i=1}^N\sum_{t=1}^T \left[y_{it} - \mathbf{x}_{it}'\beta - \alpha_i - a - b\left(\frac{t}{T} - \tau\right)\right]^2 k\left(\frac{t-\tau T}{Th}\right)$$

for given $\alpha = (\alpha_2, \ldots, \alpha_N)'$ and β. Then, α and β can be estimated by

$$(\hat{\alpha}', \hat{\beta}')' = \arg\min_{(\alpha',\beta')'} \sum_{i=1}^N\sum_{t=1}^T \left[y_{it} - \mathbf{x}_{it}'\beta - \alpha_i - \hat{f}_{\alpha,\beta}(\frac{t}{T})\right]^2.$$

For the purpose of deriving the asymptotic properties of $\hat{\beta}$ and $\hat{f}(\tau)$, Chen et al. (2012) impose the following moment conditions on $\{\mathbf{v}_{it}\}$, which allow for cross-sectional dependence,

$$\frac{1}{N}\sum_{i=1}^N E(\mathbf{v}_{it}\mathbf{v}_{it}') \to \Sigma_v, \quad \sum_{i=1}^N\sum_{j=1}^N \sigma_v(i,j) = O(N), \quad \text{and} \quad E\left\|\sum_{i=1}^N \mathbf{v}_{it}\right\|^\delta = O(N^{\delta/2})$$

as $N \to \infty$, where Σ_v is a positive definite matrix and $\sigma_v(i,j) = E(\mathbf{v}_{i1}\mathbf{v}_{j1}') + 2\sum_{t=2}^\infty E(\mathbf{v}_{i1}\mathbf{v}_{jt}')$.

Similar conditions are also imposed on $\{e_{it}\}$. Under fairly general conditions, Chen et al. (2012) show that $\hat{\beta}$ is asymptotic normality and achieves $\sqrt{NT}$ convergence rate as both N and T tend to infinity, while the local linear estimator of the trend function is also asymptotically normally distributed with a $\sqrt{NTh}$ convergence rate.

§5 Conclusion

Since the pioneer papers by Pesaran (2006) and Bai (2009) on panel data models with cross-sectional dependence, both linear and nonlinear (nonparametric and semiparametric) panel data modeling with cross-sectional dependence has become an integral part of research in econometrics. The literature is already vast and continues to grow swiftly, involving a full spread of participants for both econometricians and statisticians and engaging a wide sweep of academic journals, including some top economics journals. The field has left indelible mark on almost all core areas in econometrics. The popularity of this field is also witnessed by the fact that graduate students and young researchers in economics, finance, mathematics, and statistics are expected to take courses in this discipline or the like and review the important research papers in this area to search for their own research interests, particularly dissertation topics for doctoral students. On the other hand, this area also has made an impact in the applied economics and financial economics. We hope that this selective review has provided the reader a perspective on this important field in econometrics and statistics and some open research problems.

Finally, we would like to point out some interesting and challenging future research topics in this field. For panel data models with cross-sectional dependence, it assumes commonly that T is large so that for each individual, regressors $\mathbf{x}_{it}$ may be possibly nonstationary such as integrated processes and trending stationary as in (22) considered by Chen at al. (2012). Therefore, a variety of panel data models with cross-sectional dependence when regressors are integrated are still open. These can be regraded an extension of the papers for functional coefficient time series models by Cai et al. (2009) and Sun et al. (2013) and time varying coefficient models in Cai and Wang (2014) and Cai at al. (2015) when regressors are integrated. Secondly, although there are some testing procedures available in the literature on testing whether the cross-sectional independence holds for real applications, as mentioned earlier, the misspecification test issues as in Su and Zhang (2013) and Cai at al. (2016) are of great importance in economics when panel models are applied to the real applications. Therefore, more research on hypothesis testing is in demand. Finally, the other type models for panel data with cross-sectional dependence such as quantile regression models as in Cai et al. (2015) are deserved for an investigation in the future.

References

[1] S C Ahn, Y H Lee, P Schmidt. *GMM estimation of linear panel data models with time-varying individual effects*, J Econometrics, 2001, 101: 219-255.

[2] L Anselin. *Spatial econometrics: methods and models*, Kluwer, Dordrecht, 1988.

[3] L Anselin, J Le Gallo, H Jayet. *Spatial panel econometrics*, In: *The Econometrics of Panel Data*, L Mátyás, P Sevestre, eds, Springer-Verlag, 2008, 625-660.

[4] J Bai. *Inferential theory for factor models of large dimensions*, Econometrica, 2003, 71: 135-171.

[5] J Bai. *Panel data models with interactive fixed effects*, Econometrica, 2009, 77: 1229-1279.

[6] B Baltagi. *Econometric analysis of panel data*, John Wiley & Sons, 2013.

[7] C A Bester, T G Conley, C B Hansen, T J Vogelsang. *Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators*, Econometric Theory, 2016, 32: 154-186.

[8] Z Cai, L Chen, Y Fang. *Quantile panel data models with partially varying coefficients with an application to the growth effect of FDI*, submitted, 2015.

[9] Z Cai, Y Fang, Q Xu. *Inferences for varying-coefficient panel data models with cross-sectional dependence*, working paper, Xiamen Univ, 2016.

[10] Z Cai, Q Li. *Nonparametric estimation of varying coefficient dynamic panel data models*, Econometric Theory, 2008, 24: 1321-1342.

[11] Z Cai, Q Li, Y Park. *Functional-coefficient models for nonstationary time series data*, J Econometrics, 2009, 148: 101-113.

[12] Z Cai, Y Wang. *Testing predictive regression models with nonstationary regressors*, J Econometrics, 2014, 178: 4-14.

[13] Z Cai, Y Wang, Y Wang. *Testing instability in predictive regression model with nonstationary regressors*, Econometric Theory, 2015, 31: 953-980.

[14] C Castagnetti, E Rossi, L Trapani. *Inference on factor structures in heterogeneous panels*, J Econometrics, 2015, 184: 145-157.

[15] J Chen, J Gao, D Li. *Semiparametric trending panel data models with cross-sectional dependence*, J Econometrics, 2012, 171: 71-85.

[16] A Chudik, M H Pesaran. *Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors*, J Econometrics, 2015, 188: 393-420.

[17] A Chudik, M H Pesaran, E Tosetti. *Weak and strong cross-section dependence and estimation of large panels*, Econom J, 2011, 14: C45-C90.

[18] T G Conley. *GMM estimation with cross sectional dependence*, J Econometrics, 1999, 92: 1-45.

[19] T G Conley, G Topa. *Socio-economic distance and spatial patterns in unemployment*, J Appl Econometrics, 2002, 17: 303-327.

[20] M Deistler, B D Anderson, A Filler, C Zinner, W Chen. *Generalized linear dynamic factor models: An approach via singular autoregressions*, Eur J Control, 2010, 16: 211-224.

[21] C Dong, J Gao, B Peng. *Semiparametric single-index panel data models with cross-sectional dependence*, J Econometrics, 2015, 188: 301-312.

[22] J P Elhorst. *Specification and estimation of spatial panel data models*, Int Regional Sci Rev, 2003, 26: 244-268.

[23] J P Elhorst. *Spatial Panel Data Models*, in *Handbook of Applied Spatial Analysis*, M M Fischer, A Getis, eds, Springer-Verlag, 2010, 377-407.

[24] G Everaert, T D Groote. *Common correlated effects estimation of dynamic panels with cross-sectional dependence*, Econometric Rev, 2016, 35: 428-463.

[25] M Forni, M Lippi. *The generalized dynamic factor model: representation theory*, Econometric Theory, 2001, 17: 1113-1141.

[26] R Greenaway-McGrevy, C Han, D Sul. *Asymptotic distribution of factor augmented estimators for panel regression*, J Econometrics, 2012, 169: 48-53.

[27] M Harding, C Lamarche. *Least squares estimation of a panel data model with multifactor error structure and endogenous covariates*, Econom Lett, 2011, 111: 197-199.

[28] D J Henderson, R J Carroll, Q Li. *Nonparametric estimation and testing of fixed effects panel data models*, J Econometrics, 2008, 144: 257-275.

[29] C Hsiao. *Analysis of Panel Data*, Cambridge Univ Press, 2014.

[30] C Hsiao, A K Tahmiscioglu. *Estimation of dynamic panel data models with both individual and time-specific effects*, J Statist Plann Inference, 2008, 138: 2698-2721.

[31] X Huang. *Nonparametric estimation in large panels with cross-sectional dependence*, Econometric Rev, 2013, 32: 754-777.

[32] G Kapetanios, M H Pesaran, T Yamagata. *Panels with non-stationary multifactor error structures*, J Econometrics, 2011, 160: 326-348.

[33] M Kapoor, H H Kelejian, I R Prucha. *Panel data models with spatially correlated error components*, J Econometrics, 2007, 140: 97-130.

[34] H Karabiyik, J-P Urbain, J Westerlund. *CCE estimation of factor-augmented regression models with more factors than observables*, working paper, Maastricht Univ, 2014.

[35] H H Kelejian, I R Prucha. *A generalized moments estimator for the autoregressive parameter in a spatial model*, Internat Econom Rev, 1999, 40: 509-533.

[36] N M Kiefer, T J Vogelsang. *A new asymptotic theory for heteroskedasticity-autocorrelation robust tests*, Econometric Theory, 2005, 21: 1130-1164.

[37] L Lee. *GMM and 2SLS estimation of mixed regressive, spatial autoregressive models*, J Econometrics, 2007, 137: 489-514.

[38] L Lee, J Yu. *Estimation of spatial autoregressive panel data models with fixed effects*, J Econometrics, 2010a, 154: 165-185.

[39] L Lee, J Yu. *A spatial dynamic panel data model with both time and individual fixed effects*, Econometric Theory, 2010b, 26: 564-597.

[40] L Lee, J Yu. *Some recent developments in spatial panel data models*, Reg Sci Urban Econ, 2010c, 40: 255-271.

[41] L Lee, J Yu. *Efficient GMM estimation of spatial dynamic panel data models with fixed effects*, J Econometrics, 2014, 180: 174-197.

[42] Z Lin, Q Li, Y Sun. *A consistent nonparametric test of parametric regression functional form in fixed effects panel data models*, J Econometrics, 2014, 178: 167-179.

[43] H R Moon, M Weidner. *Dynamic linear panel regression models with interactive fixed effects*, USC-INET Research Paper No 15-02, 2014.

[44] H R Moon, M Weidner. *Linear regression for panel with unknown number of factors as interactive fixed effects*, Econometrica, 2015, 83: 1543-1579.

[45] M H Pesaran. *Estimation and inference in large heterogeneous panels with a multifactor error structure*, Econometrica, 2006, 74: 967-1012.

[46] M H Pesaran, E Tosetti. *Large panels with common factors and spatial correlation*, J Econometrics, 2011, 161: 182-202.

[47] P C Phillips, D Sul. *Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence*, J Econometrics, 2007, 137: 162-188.

[48] P M Robinson. *Nonparametric trending regression with cross-sectional dependence*, J Econometrics, 2012, 169: 4-14.

[49] V Sarafidis, T Wansbeek. *Cross-sectional dependence in panel data analysis*, Econometric Rev, 2012, 31: 483-531.

[50] M Song. *Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications*, working paper, Columbia Univ, 2013.

[51] L Su, S Jin. *Sieve estimation of panel data models with cross section dependence*, J Econometrics, 2012, 169: 34-47.

[52] L Su, S Jin, Y Zhang. *Specification test for panel data models with interactive fixed effects*, J Econometrics, 2015, 186: 222-244.

[53] L Su, Z Yang. *QML estimation of dynamic panel data models with spatial errors*, J Econometrics, 2015, 185: 230-258.

[54] L Su, Y Zhang. *Nonparametric dynamic panel data models with interactive fixed effects: sieve estimation and specification testing*, working paper, Singapore Management Univ, 2013.

[55] Y Sun, Z Cai, Q Li. *Semiparametric functional coefficient models with integrated covariates*, Econometric Theory, 2013, 29: 659-672.

[56] W R Tobler. *A computer movie simulating urban growth in the detroit region*, Econ Geogr, 1970, 46: 234-240.

[57] J Westerlund, J-P Urbain. *On the estimation and inference in factor-augmented panel regressions with correlated loadings*, Econom Lett, 2013, 119: 247-250.

[58] J Westerlund, J-P Urbain. *Cross-sectional averages versus principal components*, J Econometrics, 2015, 185: 372-377.

[59] J Yu, R de Jong, L Lee. *Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large*, J Econometrics, 2008, 146: 118-134.

[60] A Zellner. *An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias*, J Amer Statist Assoc, 1962, 57: 348-368.

[1] The Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics (Xiamen University), Fujian Provincial Key Laboratory of Statistical Science, Xiamen University, Xiamen 361005, China. Email: yifst1@gmail.com

[2] Department of Economics, University of Kansas, Lawrence, KS 66045, USA.